In [89]:

```
# ============================================================
# Assignment: Final Project - Phase 1
# Filename: f1.py
# Description: Final Project
# Date: 04/14/2020
# Author: Tarini Dash
# ============================================================

import pandas as pd
import math
import matplotlib.pyplot as plt


def main():
    fig, axs = plt.subplots(9,figsize=(15,100))
    count = 0
    col = ["Scn", "A2", "A3", "A4", "A5", "A6", "A7",
```

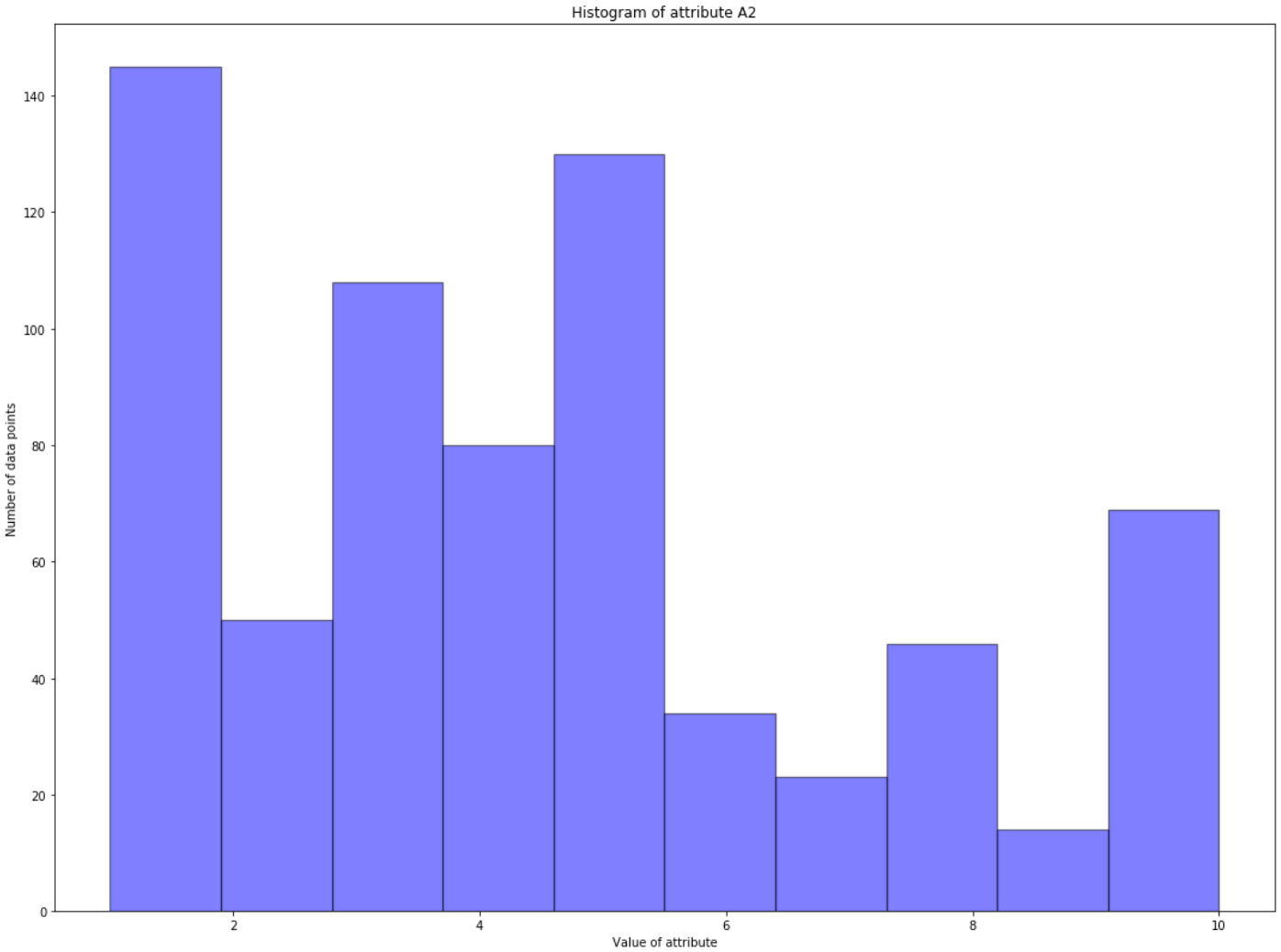
```
"A8", "A9", "A10", "Class"]

    # read from file
    df = pd.read_csv('breast-cancer-wisconsin.data',
na_values = '?', names = col)

    # drop columns
    df = df.drop(columns=['Scn', 'Class'])

    # replace missing value with mean
    df = df.fillna(df.mean()['A7'])


    for col_name in df.columns:
        print("Attribute",col_name," ----------------
----------")
        print("\t Mean \t\t\t:", round(df[col_name].m
ean(),1))
        print("\t Median \t\t:", round(df[col_name].m
edian(),1))
        print("\t Variance \t\t:", round(df[col_name]
```

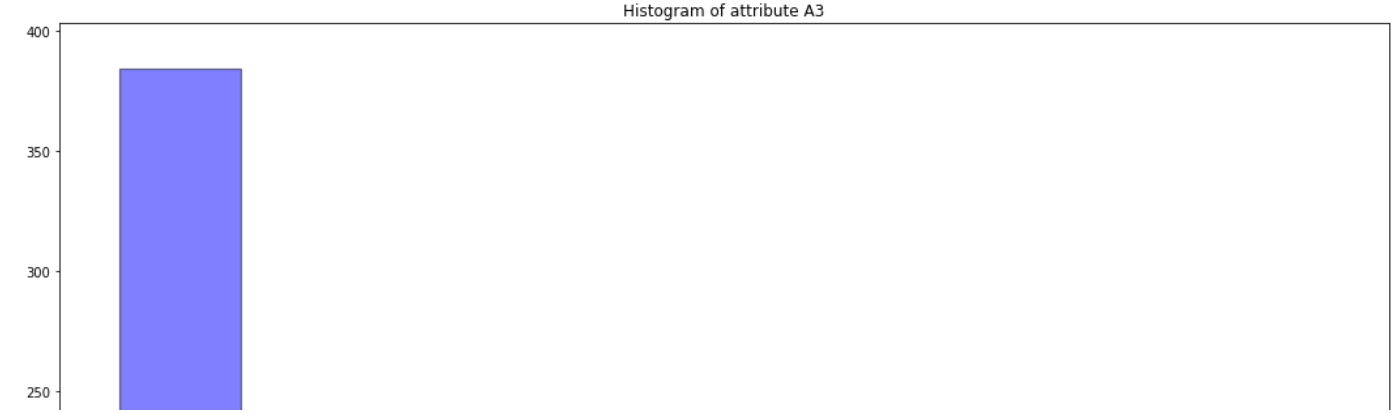
```
.var(),1))
        print("\t Standard Deviation \t:", round(df[c
ol_name].std(),1), end="\n\n")
        axs[count].set_title("Histogram of attribute
 " + col_name)
        axs[count].set_xlabel("Value of attribute")
        axs[count].set_ylabel("Number of data points"
)
        axs[count].hist(df[col_name],bins=10, color=
"blue", edgecolor='black',linewidth=1.2, alpha=0.5)
        count += 1

    fig.tight_layout()


#invoke main method
if __name__ == "__main__":
    main()
```

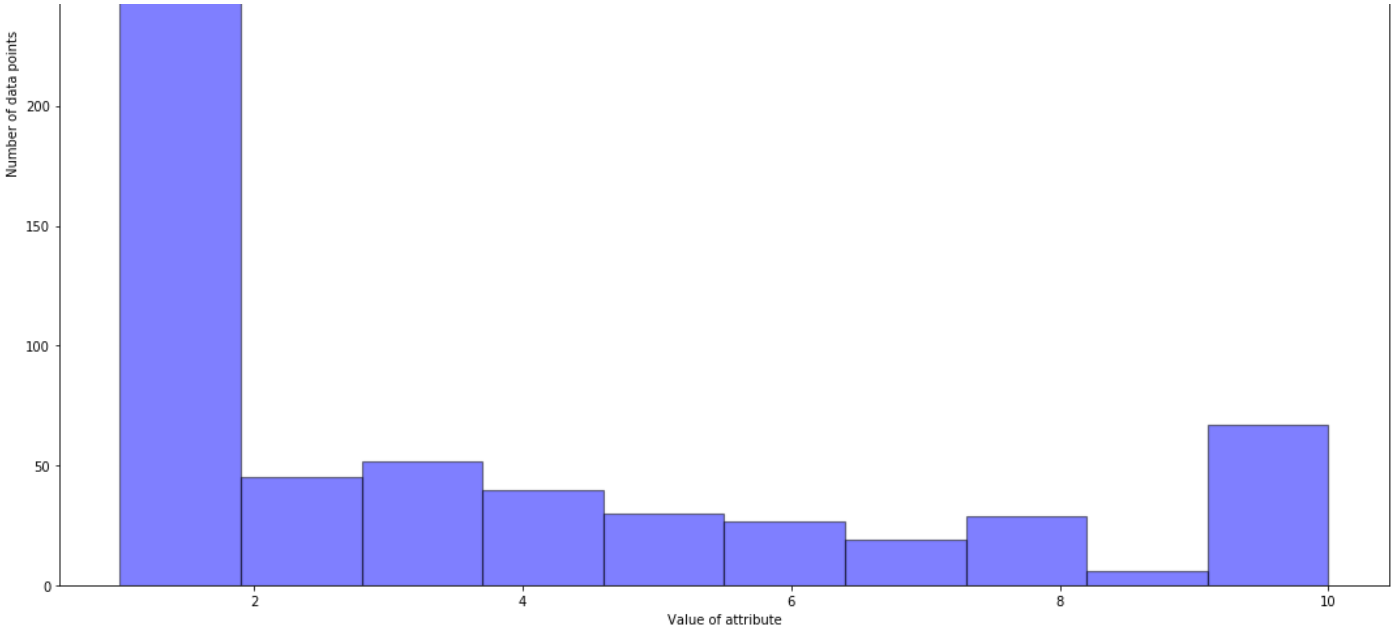
```
Attribute A2  ----------------------------
          Mean                        : 4.4
          Median                      : 4.0
          Variance                    : 7.9
          Standard Deviation          : 2.8

Attribute A3  ----------------------------
          Mean                        : 3.1
          Median                      : 1.0
          Variance                    : 9.3
          Standard Deviation          : 3.1

Attribute A4  ----------------------------
          Mean                        : 3.2
          Median                      : 1.0
          Variance                    : 8.8
          Standard Deviation          : 3.0

Attribute A5  ----------------------------
          Mean                        : 2.8
          Median                      : 1.0
```

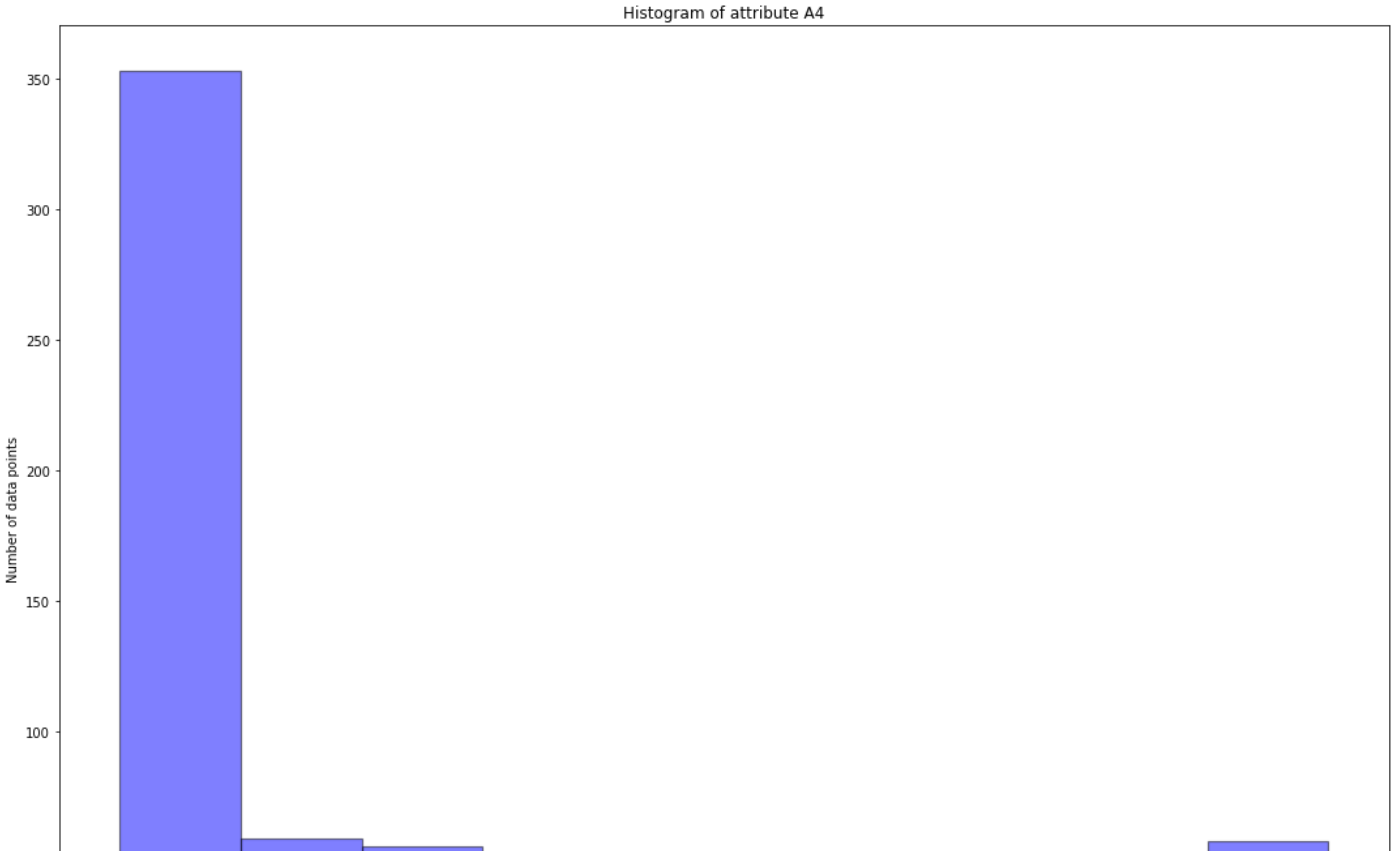
```
          Variance                  : 8.2
          Standard Deviation        : 2.9

Attribute A6  --------------------------
          Mean                      : 3.2
          Median                    : 2.0
          Variance                  : 4.9
          Standard Deviation        : 2.2

Attribute A7  --------------------------
          Mean                      : 3.5
          Median                    : 1.0
          Variance                  : 13.0
          Standard Deviation        : 3.6

Attribute A8  --------------------------
          Mean                      : 3.4
          Median                    : 3.0
          Variance                  : 5.9
          Standard Deviation        : 2.4
```

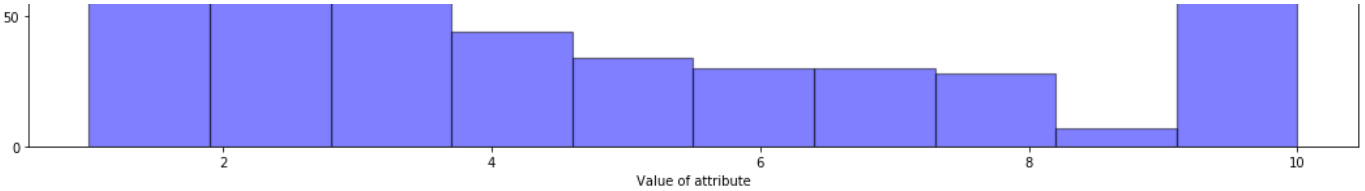
```
Attribute A9  ----------------------------
          Mean                       : 2.9
          Median                     : 1.0
          Variance                   : 9.3
          Standard Deviation         : 3.1

Attribute A10  ---------------------------
          Mean                       : 1.6
          Median                     : 1.0
          Variance                   : 2.9
          Standard Deviation         : 1.7
```

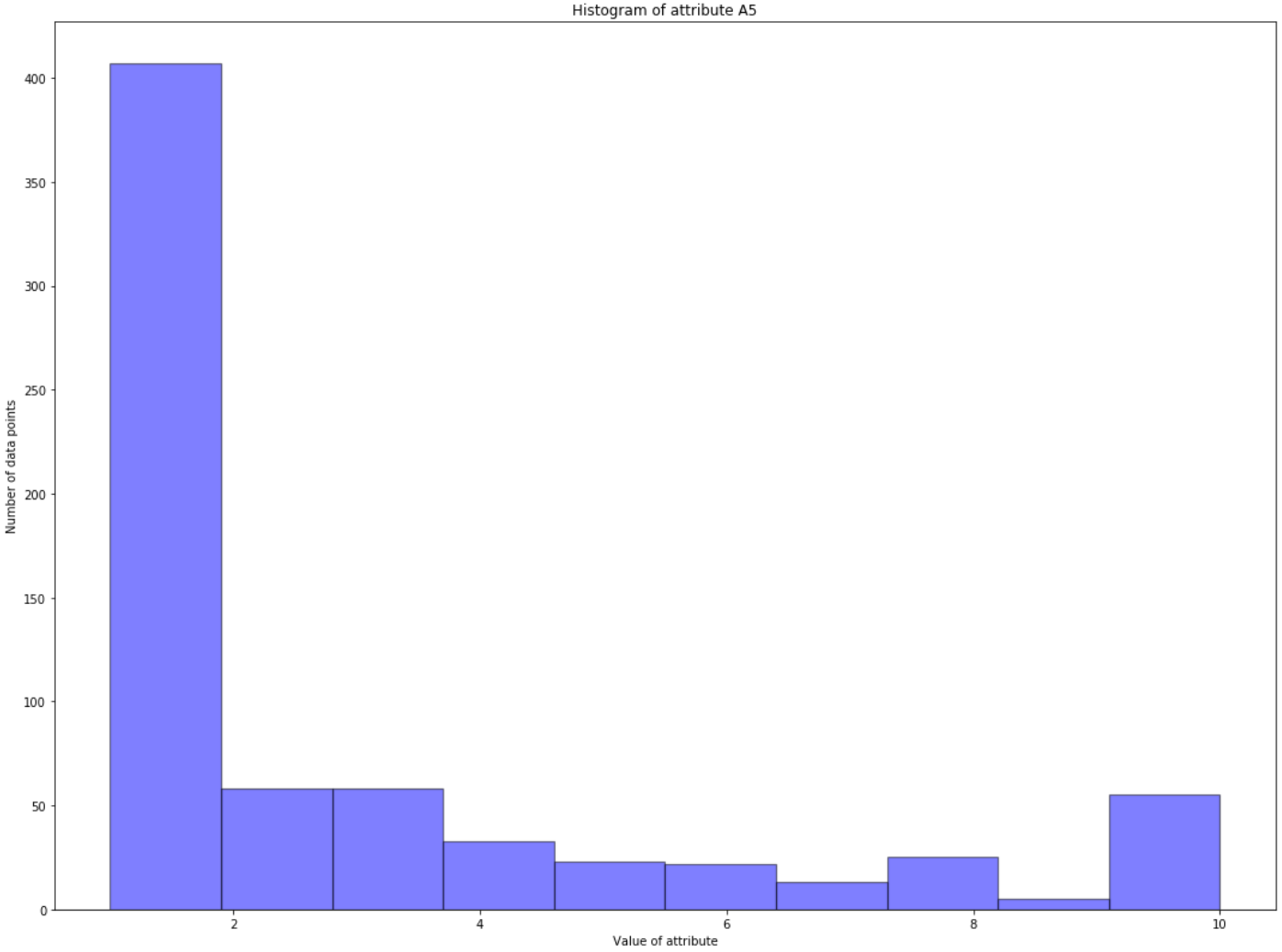

Histogram of attribute A2
Number of data points
Value of attribute
0
20
40
60
80
100
120
140
2
4
6
8
10

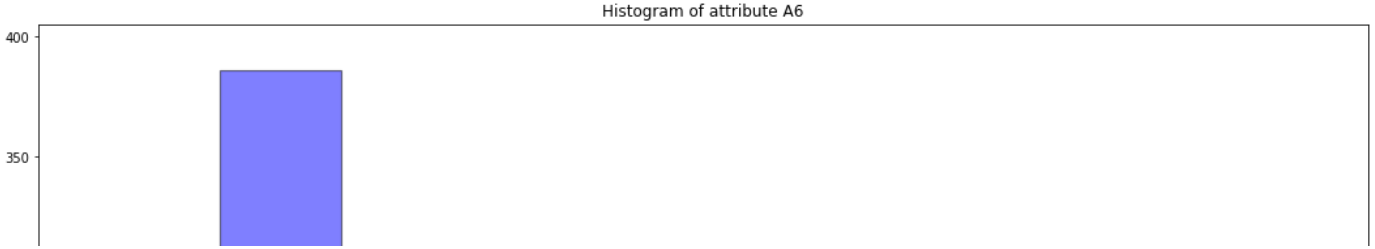

Histogram of attribute A3
250
300
350
400

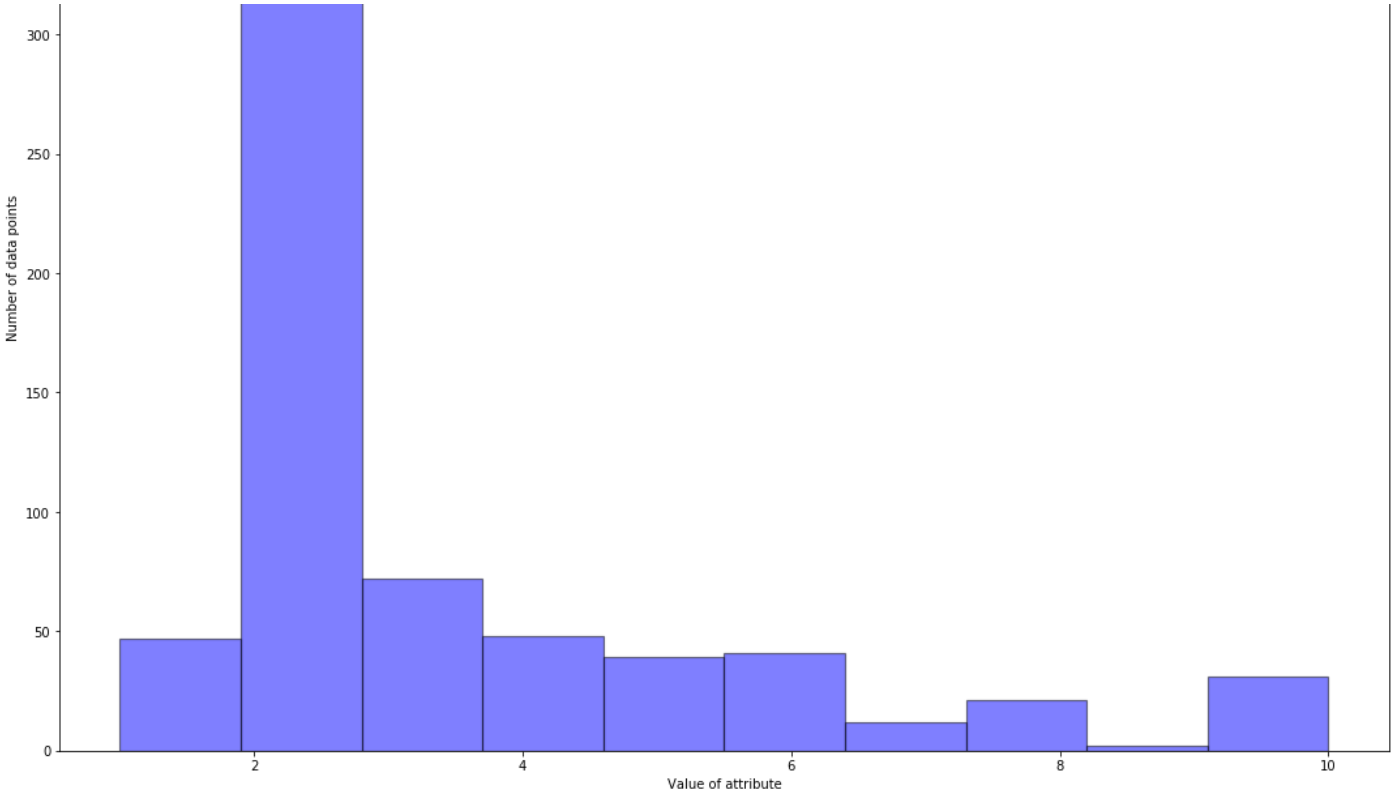

Number of data points
0
50
100
150
200
2
4
6
8
10
Value of attribute


Histogram of attribute A4
Number of data points
100
150
200
250
300
350

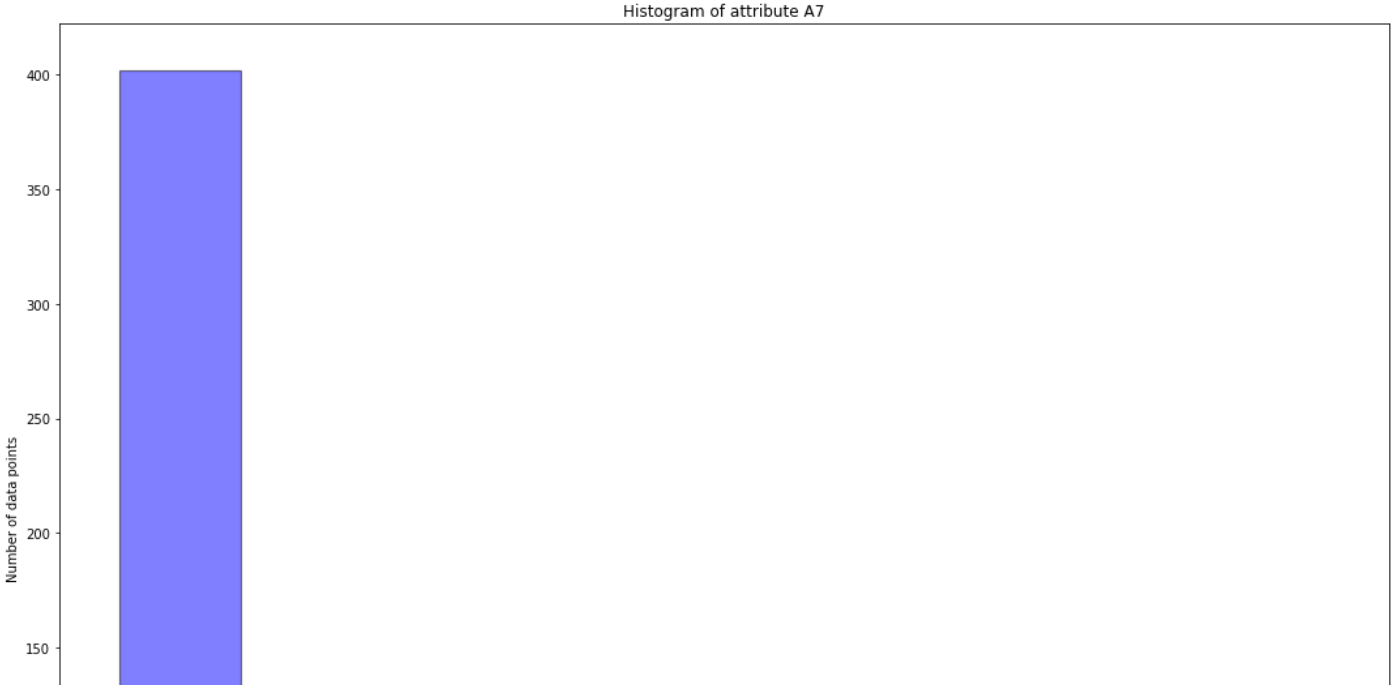

50
0
2
4
6
8
10
Value of attribute

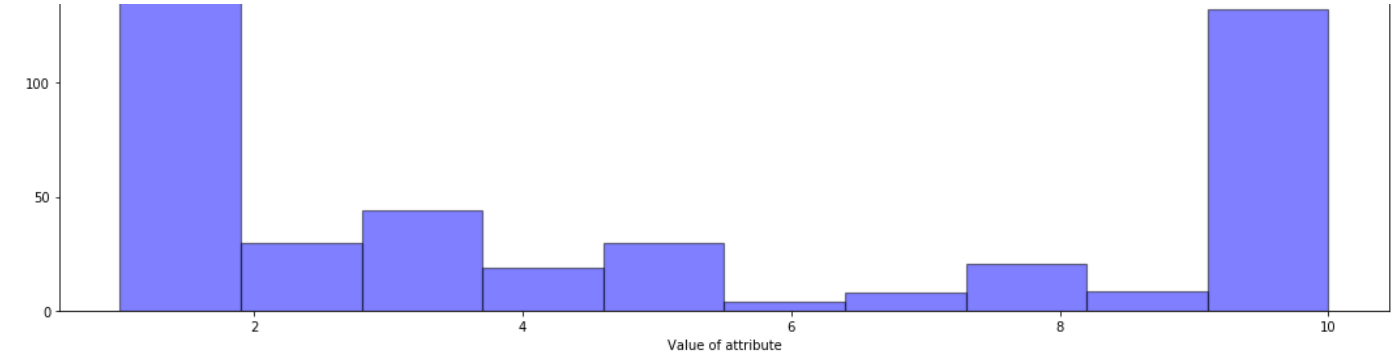

Histogram of attribute A5
400
350
300
250
200
150
100
50
0
Number of data points
2
4
6
8
10
Value of attribute


Histogram of attribute A6
400
350

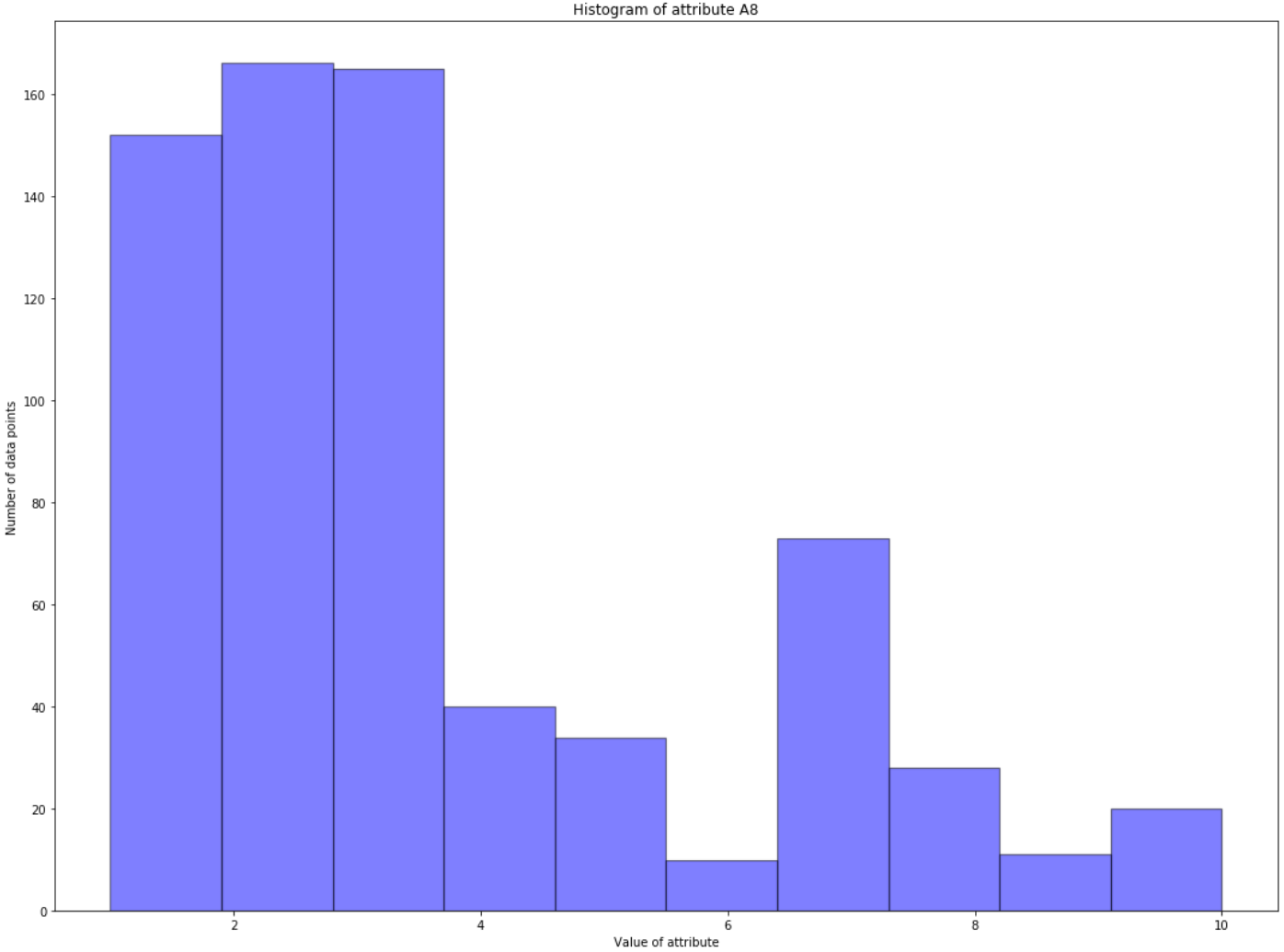


Histogram of attribute A7

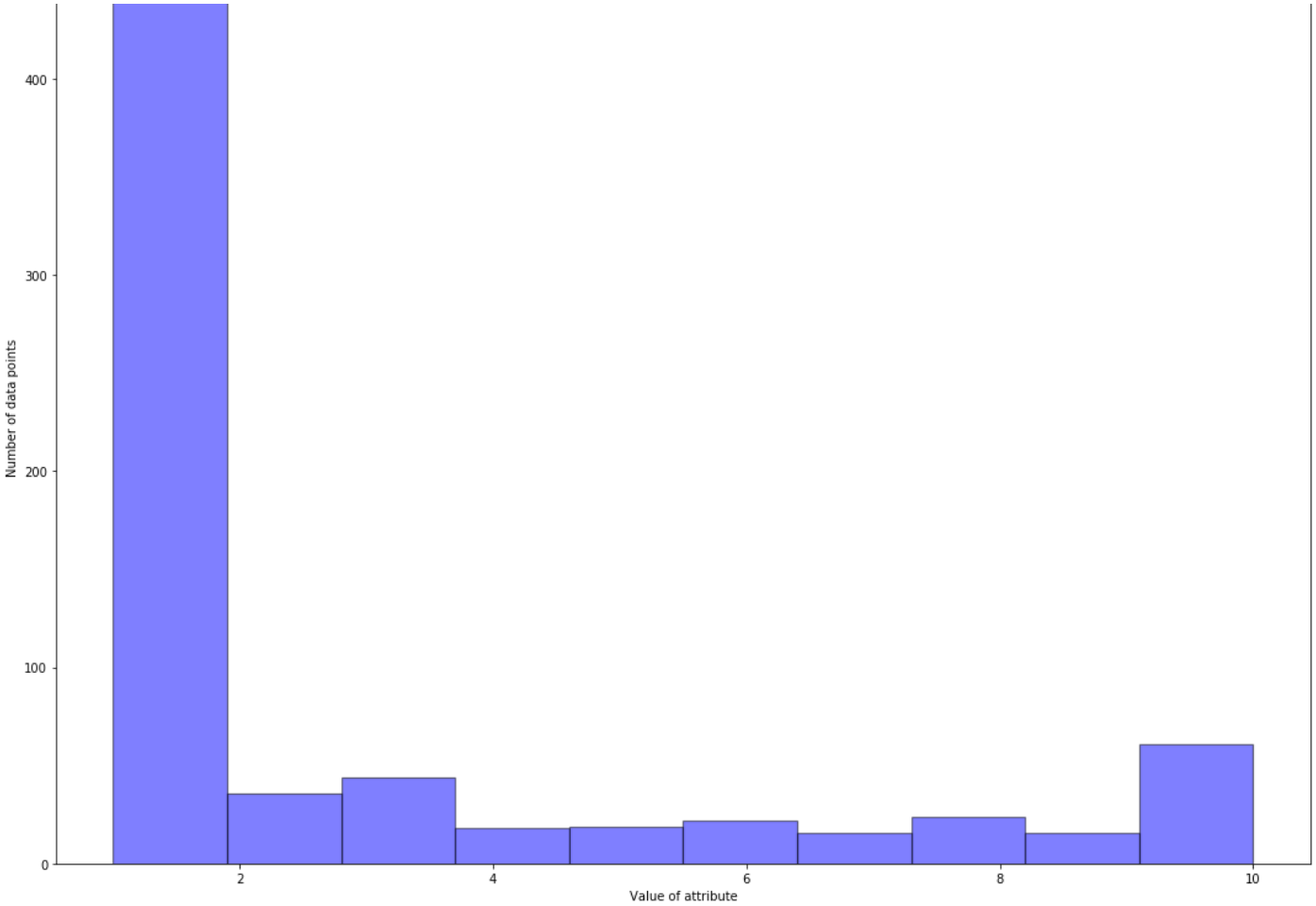




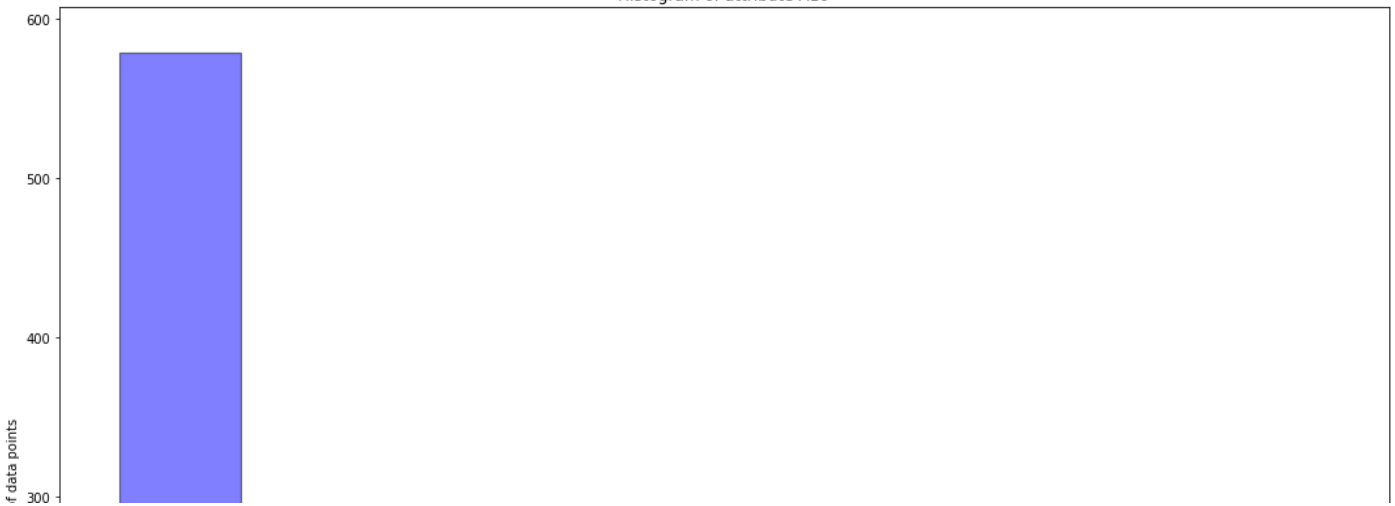
Histogram of attribute A8

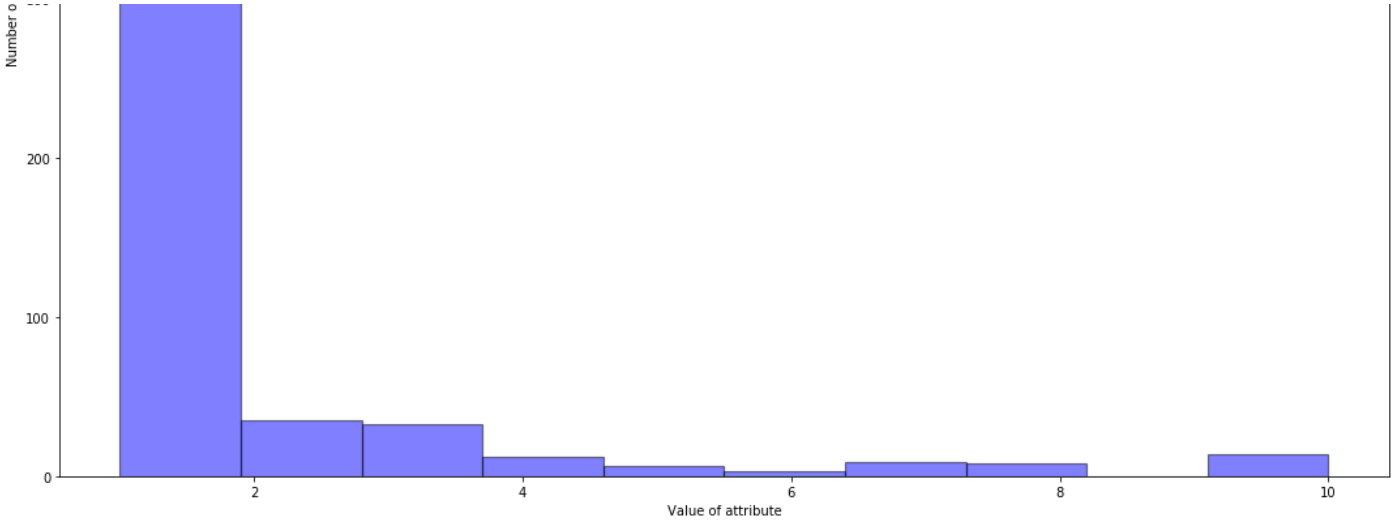


Histogram of attribute A9



Histogram of attribute A10





In [ ]: